(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2001－101184
（P2001－101184A）
(43)公開日　平成13年４月13日(2001.4.13)

(51)Int.Cl.'	識別記号	ＦＩ		テーマコード'(参考)
Ｇ０６Ｆ　17/27		Ｇ０６Ｆ　15/20	５５０Ｅ	５Ｂ００９
17/30		15/40	３４０	５Ｂ０７５
			３７０Ａ	
		15/401	３２０Ａ	

審査請求　未請求　請求項の数６　ＯＬ　（全８頁）

(21)出願番号　特願平11－281937

(22)出願日　平成11年10月１日(1999.10.1)

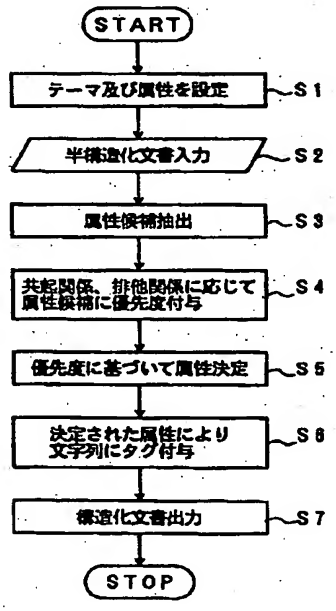
(71)出願人　000004226
日本電信電話株式会社
東京都千代田区大手町二丁目３番１号
(72)発明者　井上　香織
東京都千代田区大手町二丁目３番１号　日本電信電話株式会社内
(72)発明者　横路　誠司
東京都千代田区大手町二丁目３番１号　日本電信電話株式会社内
(74)代理人　100070150
弁理士　伊東　忠彦

最終頁に続く

(54)【発明の名称】　構造化文書生成方法及び装置及び構造化文書生成プログラムを格納した記憶媒体

(57)【要約】　（修正有）
【課題】　非構造化文書を構造化する際の属性判定基準をテーマによって可変とし、検索時に検索者が選択したテーマ別の検索を可能とする。
【解決手段】　各テーマ毎に基本的な属性セットを設定しておき、半構造化文書が入力されると、該文書の文字列に対して予め登録されているパターンとのパターンマッチング、及び単語と複数の属性名が対応して記述されている属性辞書との辞書マッチングを行って、該半構造化文書の文字列に対する属性候補を抽出し、テーマ毎の属性を参照して、半構造化文書中に出現する可能性のある属性を取得すると共に、属性同士が共起関係にあるか、排他関係にあるかを示す属性関係ルールを参照して優先度を付与し、属性候補のうち、優先度が大きいものを属性として採用し、構造化文書として出力する。

本発明の原理を説明するための図



1

【特許請求の範囲】

【請求項１】　テーマ別文書検索を目的として、テーマに基づいた非構造化文書の構造化を行うための構造化文書生成方法において、

予め検索の視点であるテーマを設定すると共に、各テーマ毎に基本的な属性セットを設定しておき、

半構造化文書が入力されると、該文書の文字列に対して予め登録されているパターンとのパターンマッチング、及び単語と複数の属性名が対応して記述さている属性辞書との辞書マッチングを行って、該半構造化文書の文字列に対する属性候補を抽出し、

抽出された前記属性候補について、テーマ毎の属性を参照して、前記半構造化文書中に出現する可能性のある属性を取得すると共に、該半構造化文書中に出現する可能性のある属性同士が共起関係にあるか、排他関係にあるかを示す属性関係ルールを参照して共起関係または、排他関係に応じて優先度を付与し、

前記属性候補のうち、前記優先度が大きいものを属性として採用し、

採用された属性に基づいて入力された前記半構造化文書の文字列に対してタグ付けを行い、構造化文書として出力することを特徴とする構造化文書生成方法。

【請求項２】　前記属性候補の優先度と所定の閾値を比較して、該優先度が該閾値を下回る場合には、該属性候補を削除する請求項１記載の構造化文書生成方法。

【請求項３】　テーマ別文書検索を目的として、テーマに基づいた非構造化文書の構造化を行うための構造化文書生成装置であって、

予め検索の視点であるテーマを指定するテーマ指定手段と、

各テーマ毎に基本的な属性セットが格納されている属性セット記憶手段と、

記号、文字列、品詞を含むパターンが格納されているパターン記憶手段と、

単語と複数の属性名が対応付けられて格納されている辞書記憶手段と、

ある属性と他の属性が共起関係にあるか、排他関係にあるかを示す属性関係ルールが格納されている属性関係ルール記憶手段と、

半構造化文書を入力する半構造化文書入力手段と、

前記半構造化文書入力手段から半構造化文書が入力されると、該文書の文字列に対して前記パターン記憶手段を参照してパターンマッチングを行い、さらに、前記辞書記憶手段を参照して辞書マッチングを行い、半構造化文書の文字列に対する属性候補を抽出する属性候補抽出手段と、

前記属性候補抽出手段において抽出された前記属性候補について、前記属性セット記憶手段を参照して、前記半構造化文書中に出現する可能性のある属性を取得すると共に、前記属性関係ルール記憶手段を参照して、該半構造化文書中に出現する可能性のある属性同士が共起関係にあるか、または、排他関係にあるかに応じて優先度を付与し、該優先度が大きい属性候補を属性として採用する属性コスト計算手段と、

採用された前記属性に基づいて、入力された前記半構造化文書の文字列に対してタグ付けを行い、構造化文書として出力する構造化文書出力手段とを有することを特徴とする構造化文書生成装置。

【請求項４】　前記属性コスト計算手段は、

前記属性候補の優先度と、所定の閾値を比較して、該優先度が該閾値を下回る場合には、該属性候補を削除する手段を含む請求項３記載の構造化文書生成装置。

【請求項５】　テーマ別文書検索を目的として、テーマに基づいた非構造化文書の構造化を行うための構造化文書生成プログラムを格納した記憶媒体であって、

半構造化文書を入力させる半構造化文書入力プロセスと、

前記半構造化文書が入力されると、該文書の文字列に対して予め登録されている記号、文字列、品詞を含むパターンを参照してパターンマッチングを行い、さらに、予め単語と複数の属性名が対応付けられて登録されている辞書を参照して辞書マッチングを行い、該半構造化文書の文字列に対する属性候補を抽出する属性候補抽出プロセスと、

前記属性候補抽出プロセスにおいて抽出された前記属性候補について、各テーマ毎に予め登録されている基本的な属性セットを参照して、前記半構造化文書中に出現する可能性のある属性を取得すると共に、該半構造化文書中に出現する可能性のある属性同士が共起関係にあるか、排他関係にあるかを示す属性関係ルールを参照して、該属性候補の共起関係または、排他関係に応じて優先度を付与し、該優先度が大きい属性候補を属性として採用する属性コスト計算プロセスと、

採用された属性に基づいて入力された前記半構造化文書の文字列に対してタグ付けを行い、構造化文書として出力させる構造化文書出力プロセスとを有することを特徴とする構造化文書生成プログラムを格納した記憶媒体。

【請求項６】　前記属性コスト計算プロセスは、

前記属性候補の優先度と、所定の閾値を比較して、該優先度が該閾値を下回る場合には、該属性候補を削除するプロセスを含む請求項５記載の構造化文書生成プログラムを格納した記憶媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、構造化文書生成方法及び装置及び構造化文書生成プログラムを格納した記憶媒体に係り、特に、テーマ別文書検索を目的として、テーマに基づいた非構造化文書の構造化を行うための構造化文書生成方法及び装置及び構造化文書生成プログラムを格納した記憶媒体に関する。

3

【０００２】

【従来の技術】プレーンな文書中の情報（特性の文字列）に対して、属性を与え、その属性間の関係を明らかにすることを構造化という。属性には、意味属性と論理属性があるが、ここで扱うのは意味属性である。意味属性の場合、そのテーマ（視点）により、ある文字列に与えられる属性は異なる。

【０００３】例えば、次の２つの文はいずれも『カメラ』に関する文である。

①　“カメラ”で写真をとる方法

②　“カメラ”大幅値下げ

この２つの文で単語『カメラ』はそれぞれ異なる使われ方をしている。①の文では、写真をとる「手段としてのカメラ」であり、②の文では値下げの対象「商品としてのカメラ」である。このような、文脈毎の単語の使われ方を「属性」と呼ぶ。ちなみに「属性」を決めるのは、作者とは限らない。読者でもよいし、なんらかのシステムでもよい。従来の全文検索（例：“goo”http://www.goo.ne.jp）では、上記のような「属性」は考慮される。入力されたキーワードにマッチする全ての結果を返すため、無駄な検索結果が多く含まれてしまう。しかし、情報を属性によって分類しておけば、ユーザは本当に欲しい情報だけを得ることができる。ここで、テキスト中の情報（単語等）に属性を与えることを「構造化」と呼ぶ。

【０００４】従来の文書構造化は、テーマ（視点）を固定することで、ある文字列に与える属性を１つに特定している。例えば、文書中に「りんご」という文字列が現れた場合には、文書のテーマ（視点）によって、「果物」や「農産物」「おやつ」など様々な属性が付与される可能性がある。しかし、属性辞書に、「果物」とだけ記述することで「りんご」という文字列には常に「果物」という属性が与えられる。

【０００５】また、論理属性も、属性は特定されるので、論理関係ルールや、辞書などの属性値抽出ルールを用いて、特定の属性付与を行う。従来の文書構造化装置の例を図９に示す。同図に示す文書構造化装置は、半構造化文書入力部１１、属性値抽出部１２、抽出ルールデータベース１３、及び構造化文書出力部１４から構成される。

【０００６】当該文書構造化装置において、半構造化文書入力部１１において文書を入力すると、属性値抽出部１２が抽出ルールデータベース１３を参照して、ある文字列に対し、特定の属性を付与する。構造化文書出力部１４は、属性値が付与された文字列を統合した構造化文書を出力する。詳細は、特開平９－６９１０１に開示されている。

【０００７】

【発明が解決しようとする課題】しかしながら、上記従来の文書構造化装置では、属性判定基準をテーマによって可変とすることはできず、検索時において検索者が選択したテーマ別の検索を柔軟に行うことができないという問題がある。本発明は、上記の点に鑑みなされたもので、非構造化文書を構造化する際の属性判定基準をテーマによって可変とし、検索時に検索者が選択したテーマ別の検索を可能とする構造化文書生成方法及び装置及び構造化文書生成プログラムを格納した記憶媒体を提供することを目的とする。

4

【０００８】

【課題を解決するための手段】図１は、本発明の原理を説明するための図である。本発明（請求項１）は、テーマ別文書検索を目的として、テーマに基づいた非構造化文書の構造化を行うための構造化文書生成方法において、予め検索の視点であるテーマを設定すると共に、各テーマ毎に基本的な属性セットを設定しておき（ステップ１）、半構造化文書が入力されると（ステップ２）、該文書の文字列に対して予め登録されているパターンとのパターンマッチング、及び単語と複数の属性名が対応して記述さている属性辞書との辞書マッチングを行って、半構造化文書の文字列に対する属性候補を抽出し（ステップ３）、抽出された属性候補について、テーマ毎の属性を参照して、半構造化文書中に出現する可能性のある属性を取得すると共に、該半構造化文書中に出現する可能性のある属性同士が共起関係にあるか、排他関係にあるかを示す属性関係ルールを参照して、共起関係または、排他関係に応じて優先度を付与し（ステップ４）、属性候補のうち、優先度が大きいものを属性として採用し（ステップ５）、採用された属性に基づいて入力された半構造化文書の文字列に対してタグ付けを行い（ステップ６）、構造化文書として出力する（ステップ７）。

【０００９】本発明（請求項２）は、属性候補の優先度と、所定の閾値を比較して、該優先度が該閾値を下回る場合には、該属性候補を削除する。本発明（請求項３）は、テーマ別文書検索を目的として、テーマに基づいた非構造化文書の構造化を行うための構造化文書生成装置であって、予め検索の視点であるテーマを指定するテーマ指定手段２９と、各テーマ毎に基本的な属性セットが格納されている属性セット記憶手段２６と、記号、文字列、品詞を含むパターンが格納されているパターン記憶手段２３と、単語と複数の属性名が対応付けられて格納されている辞書記憶手段２４と、ある属性と他の属性が共起関係にあるか、排他関係にあるかを示す属性関係ルールが格納されている属性関係ルール記憶手段２７と、半構造化文書を入力する半構造化文書入力手段２１と、半構造化文書入力手段２１から半構造化文書が入力されると、該文書の文字列に対してパターン記憶手段２３を参照してパターンマッチングを行い、さらに、辞書記憶手段２３を参照して辞書マッチングを行い、該半構造化文の文字列に対する属性候補を抽出する属性候補抽出手

段２２と、属性候補抽出手段２２において抽出された属性候補について、属性セット記憶手段２６を参照して、該半構造化文書中に出現する可能性がある属性を取得すると共に、属性関係ルール記憶手段２７を参照して、該半構造化文書中に出現する可能性のある属性同士が共起関係にあるか、または、排他関係にあるかに応じて優先度を付与し、該優先度が大きい属性候補を属性として採用する属性コスト計算手段２５と、採用された属性に基づいて入力された半構造化文書の文字列に対してタグ付けを行い、構造化文書として出力する構造化文書出力手段２８とを有する。

【００１０】本発明（請求項４）属性コスト計算手段２５において、属性候補の優先度と、所定の閾値を比較して、該優先度が該閾値を下回る場合には、該属性候補を削除する手段を含む。本発明（請求項５）は、テーマ別文書検索を目的として、テーマに基づいた非構造化文書の構造化を行うための構造化文書生成プログラムを格納した記憶媒体であって、半構造化文書を入力させる半構造化文書入力プロセスと、半構造化文書が入力されると、該文書の文字列に対して予め登録されている記号、文字列、品詞を含むパターンを参照してパターンマッチングを行い、さらに、予め単語と複数の属性名が対応付けられて登録されている辞書を参照して辞書マッチングを行い、該半構造化文書の文字列に対する属性候補を抽出する属性候補抽出プロセスと、属性候補抽出プロセスにおいて抽出された属性候補について、各テーマ毎に予め登録されている基本的な属性セットを参照して、該半構造化文書中に出現する可能性のある属性を取得すると共に、該半構造化文書中に出現する可能性のある属性同士が共起関係にあるか、または、排他関係にあるかを示す属性関係ルールを参照して、該属性候補の共起関係または、排他関係に応じて優先度を付与し、該優先度が大きい属性候補を属性として採用する属性コスト計算プロセスと、採用された属性に基づいて入力された半構造化文書の文字列に対してタグ付けを行い、構造化文書として出力させる構造化文書出力プロセスとを有する。

【００１１】本発明（請求項６）は、属性コスト計算プロセスにおいて、属性候補の優先度と、所定の閾値を比較して、該優先度が該閾値を下回る場合には、該属性候補を削除するプロセスを含む。上記のように、本発明では、予め設定れたテーマ毎に基本属性セット及び基本属性関係ルールを設定することにより、属性判定基準を可変とすることが可能となる。

【００１２】さらに、文書内の情報（文字列）に初めに複数の属性を付与し属性候補としておき、基本属性の共起や排他の関係ルールを参照することにより属性を特定することが可能となり、検索時に検索者が選択したテーマ別の検索を可能とする。

【００１３】

【発明の実施の形態】図３は、本発明の文書構造化装置

の構成を示す。同図に示す文書構造化装置は、半構造化文書入力部２１、属性候補抽出部２２、パターンデータベース２３、辞書データベース２４、属性コスト計算部２５、属性セットデータベース２６、属性関係ルールデータベース２７、構造化文書出力部２８及びテーマ指定部２９から構成される。

【００１４】半構造化文書入力部２１は、半構造化文書を入力し、属性候補抽出部２２に転送する。属性候補抽出部２２は、入力された半構造化文書のある文字列について、パターンデータベース２３や辞書データベース２４を参照してパターンマッチ、及び辞書マッチを行うものであり、パターンマッチ処理部２２１、辞書マッチ処理部２２２から構成される。

【００１５】パターンマッチ処理部２２１は、パターンデータベース２３を参照してパターンマッチを行う。当該パターンデータベース２３は、記号や文字列、品詞などのパターンと、複数属性名が対応して記述されている。辞書マッチ処理部２２２は、辞書データベース２４を参照して、辞書マッチ処理を行う。当該辞書データベース２４は、単語と複数属性名が対応して記述されている。属性候補抽出部２２において、これらのデータベース２３、２４とマッチングを行うことにより、文書データに対し、全ての属性候補が抽出される。

【００１６】属性コスト計算部２５は、各属性候補のコスト計算を行う。コスト計算は、属性セットデータベース２６を参照して、テーマ毎の属性セットを取得し、次に、属性関係ルールデータベース２７を参照して、属性セットデータベース２６から取得した属性セット中の属性間関係の共起・排他関係を調べ、重み計算を行う。属性セットデータベース２６には、テーマ名とそれに対応する属性のセットが記述されている。属性関係ルールデータベース２７には、ある属性と他のある属性が共起関係にあるか、排他関係にあるかが重みで示してある。

【００１７】属性コスト計算部２５では、ある文字列に対する複数の属性候補の中から、重みが重いものを優先して属性に採用する。また、属性候補が一つしかないもので、重みが閾値を下回った場合には、その属性を削除する。構造化文書出力部２８は、属性コスト計算部２５で特定された属性を文字列に対してタグ付けして出力する。

【００１８】テーマ指定部２９は、利用者が属性コスト計算部２５に対して決定されたテーマを入力する。図４は、本発明の文書構造化装置の動作を示すフローチャートである。

ステップ１０１）　まず、前処理として、検索のテーマを複数設定しておき、データベースとルールを予め用意しておく。データベースとして、各テーマ毎に属性を属性セットデータベース２６にセットする。記号及び数字属性抽出のためのパターンを作成し、パターンデータベース２３に設定する。文字列属性を抽出するための情報

を辞書データベース２４に設定する。属性の関係を記述した属性関係ルールを属性関係ルールデータベース２７に設定する。

【００１９】ステップ１０２）　半構造化文書入力部２１から半構造化文書を入力し、属性候補抽出部２２に転送する。

ステップ１０３）　属性候補抽出部２２のパターンマッチ処理部２２１においてパターンデータベース２３を参照して、入力された半構造化文書のある文字列についてパターンマッチ処理を行う。

【００２０】ステップ１０４）　属性候補抽出部２３の辞書マッチ処理部２２２において、文字列について、辞書データベース２４を参照して辞書マッチ処理を行い、属性候補を抽出する。なお、上記のステップ１０３とステップ１０４の処理順序は、逆であってもよい。

【００２１】ステップ１０５）　属性候補抽出部２２で抽出された属性候補（一時的なスプールに格納する）を属性コスト計算部２５に送る。テーマ指定部２９から利用者により決定されたテーマを属性コスト計算部２５に入力し、属性コスト計算部２５は、属性候補に対して、属性セットデータベース２６と属性関係ルールデータベース２７を参照して優先度を付与する。

【００２２】ステップ１０６）　属性コスト計算部２５は、属性候補に付与された優先度が所定の閾値以下の候補であるかを判定し、閾値以下である場合にはステップ１０７に移行し、閾値より大きい場合には、ステップ１０８に移行する。

ステップ１０７）　属性候補の優先度が低所定の閾値以下の場合には、当該属性候補を削除する。

【００２３】ステップ１０８）　属性候補の優先度が低所定の閾値より大きい場合には、当該属性候補の各テーマ毎の属性を決定し、元の文書（半構造化文書）にタグを付与し、構造化文書として構造化文書出力部２８から出力する。

【００２４】

【実施例】以下、図面と共に本発明の実施例を説明する。以下の説明では、前述の図４の処理に基づいて説明する。図５は、本発明の一実施例の入力される半構造化文書の例を示す。まず、図５に示す半構造化文書を半構造化文書入力部２１から入力する（ステップ１０２）。

【００２５】入力された半構造化文書に対し、属性候補抽出部２２において、属性候補抽出処理行う。パターンマッチ処理部２２１は、パターンデータベース２３を参照した結果、ＸＸ－ＸＸＸＸ－ＸＸＸＸというパターンと、「電話番号」「ＦＡＸ番号」属性が対応している場合、図５中の文字列、“０３－３３３Ｘ－０００Ｘ”とパターンがマッチするため、“０３－３３３Ｘ－０００Ｘ”には、「電話番号」と「ＦＡＸ番号」属性が候補として登録される（ステップ１０３）。

【００２６】辞書マッチ処理部２２２は、辞書データベース２４を参照して、辞書マッチ処理を行う。例えば、「ハンバーグ」という単語は「メニュー」「商品」という意味属性を持つ、と辞書データベース２４に記述されていた場合には、図５中の文字列「ハンバーグ」に対して、「メニュー」と「商品」という複数の属性が取得される（ステップ１０４）。

【００２７】属性コスト計算部２５において、各属性候補の計算を行う。属性セットデータベース２６を参照して、テーマ毎の属性セットを取得する。この例を図６に示す。例えば、図５に示す半構造化文書を「テーマ：レストラン広告」という視点で構造化したい、とし、当該テーマをテーマ指定部２９より入力する。対応する属性セットとして、「店名、メニュー、住所、…」などが取得できる（ステップ１０５）。

【００２８】ここで、ステップ１０３、１０４で抽出した属性候補の中で属性セットに含まれていなかった属性候補を削除する（ステップ１０７）。次に、属性コスト計算部２５は、属性関係ルールデータベース２７を参照して得た属性セット中の属性間関係の共起、排他関係を調べ、重み計算を行う。この例を図７に示す。図７の例では、左に、属性（属性値）の組み合わせ、右に、その重み付けコストが記述されている。コストはプラスが共起ルール、マイナスが排他ルールである。ここで、図３の例の、文字列「ハンバーグ」に対する属性候補「メニュー」と「商品名」では、「メニュー」の方が合計の重みが重いので、「ハンバーグ」に「メニュー」属性を付与する（ステップ１０５）。

【００２９】構造化文書出力部２２において、特定された属性を文字列に対してタグ付けして、出力する。例えば、図８に示すように、ＸＭＬ文書として出力する。また、上記の実施例では、図３に示す構成に基づいて説明しているが、この例に限定されることなく、半構造化文書入力部、属性候補抽出部及び属性コスト計算部及び構造化文書出力部をプログラムとして構築し、構造化文書生成装置として使用されるコンピュータに接続されるディスク装置や、フロッピーディスク、ＣＤ－ＲＯＭ等の可搬記憶媒体に格納しておき、本発明を実施する際にインストールすることにより、容易に本発明を実現することができる。

【００３０】なお、本発明は、上記の実施例に限定されることなく、特許請求の範囲内において、種々変更・応用が可能である。

【００３１】

【発明の効果】上述のように、本発明によれば、非構造化文書を構造化する際の属性判定基準をテーマによって可変とし、検索時に検索者が選択したテーマ別の検索を行うことができる。

【図面の簡単な説明】

【図１】本発明の原理を説明するための図である。

【図２】本発明の原理構成図である。

9

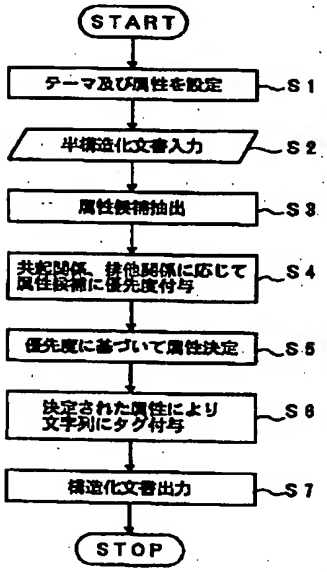
【図３】本発明の文書構造化装置の構成図である。
【図４】本発明の文書構造化装置の動作を示すフローチャートである。
【図５】本発明の一実施例の入力される半構造化文書の例である。
【図６】本発明の一実施例のテーマ別属性セットの例である。
【図７】本発明の一実施例の属性関係ルールの例である。
【図８】本発明の一実施例の出力の構造化文書の例である。
【図９】従来の文書構造化装置の構成図である。
【符号の説明】

10

＊２１　半構造化文書入力手段、半構造化文書入力部
２２　属性候補抽出手段、属性候補抽出部
２３　パターン記憶手段、パターンデータベース
２４　辞書記憶手段、辞書データベース
２５　属性コスト計算手段、属性コスト計算部
２６　属性セット記憶手段、属性セットデータベース
２７　属性関係ルール記憶手段、属性関係ルールデータベース
２８　構造化文書出力手段、構造化文書出力部
２９　テーマ指定手段、テーマ指定部
２２１　パターンマッチ処理部
２２２　辞書マッチ処理部

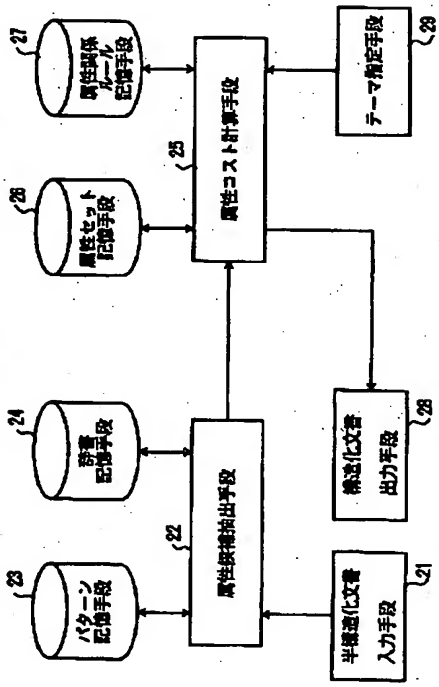
【図１】

本発明の原理を説明するための図



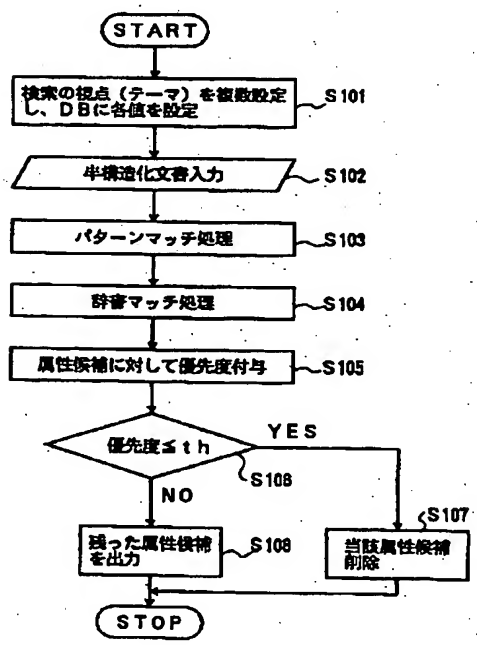
【図２】

本発明の原理構成図



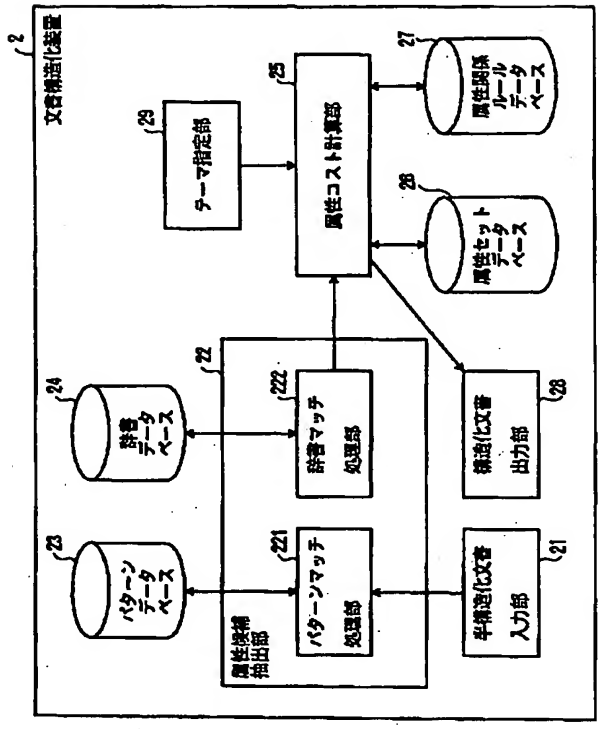
【図４】

本発明の文書構造化装置の動作を示すフローチャート



【図３】

本発明の文書構造化装置の構成図



【図５】

本発明の一実施例の入力される半構造化文書の例

レストラン あいうえお

ハンバーグ
グラタン
スパゲティ
アイスクリーム

東京都武蔵野市○○町
03-333X-000X
おいしいよ！

【図６】

本発明の一実施例のテーマ別属性セットの例

テーマ：　レストラン広告
属性セット：　店名
メニュー
値段
住所
電話番号
座席数
…
…

【図７】

本発明の一実施例の属性関係ルールの例

（業種名：レストラン、メニュー）	＋３
（業種名：スーパー、商品名）	＋３
（業種名：レストラン、商品名）	－１
…	
…	

【図８】

本発明の一実施例の出力の構造化文書の例

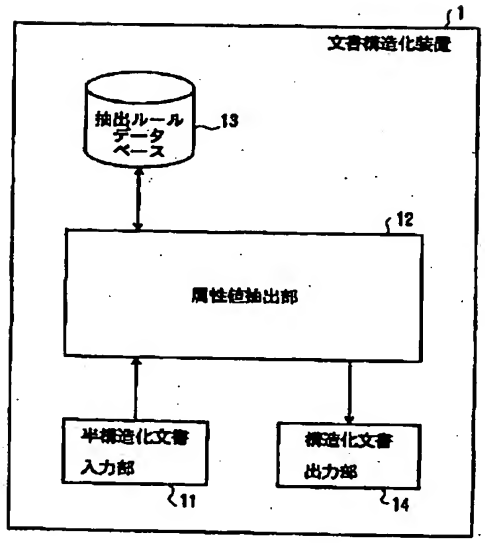
<店名>レストラン あいうえお</店名>

<メニュー>ハンバーグ</メニュー>
<メニュー>グラタン</メニュー>
<メニュー>スパゲティ</メニュー>
<メニュー>アイスクリーム</メニュー>

<住所>東京都武蔵野市○○町</住所>
<電話番号>03-333X-000X</電話番号>
<コメント>おいしいよ！</コメント>

【図９】

従来の文書構造化装置の構成図



フロントページの続き

(72)発明者　高橋　克巳
東京都千代田区大手町二丁目３番１号　日本電信電話株式会社内

Ｆターム(参考)　5B009 QA00
5B075 ND03 NK02 NK32 NK42 NK46 NR03 NR12 PP02 PP12 PP25 PR08 QM10 UU06